Converting Personal Comic Book Collection Records to Linked Data

Sean Petiya
Kent State University, USA
spetiya1@kent.edu

Keywords: Linked Data; comic books; graphic novels; ontologies; metadata; usability.

1. Introduction

The Comic Book Ontology (CBO) is a metadata vocabulary designed for the description of comic books and comic book collections. The vocabulary is part of a larger, ongoing research project exploring the design and exchange of data about comic books and graphic novels. The goal of the project is to produce a series of usable schemata and tools for the many participants in the often complex universe of comic books, which includes publishers, collectors, and libraries, among many others. The long-term objectives of the project include addressing the needs and overlapping roles of each user group through designated application profiles. Recognizing that all groups involved will have different needs, goals, and concerns, the base for each of these user application profiles is a much simpler set of elements required to first uniquely identify a resource. The intention of this core set of elements is to lower the difficulty in implementing the vocabulary and enhance the overall understandability of the ontology. The core application profile has been modeled from common elements found in the data of comic book collectors, a community of users largely responsible for the preservation of the medium, which has traditionally been underrepresented in knowledge institutions.

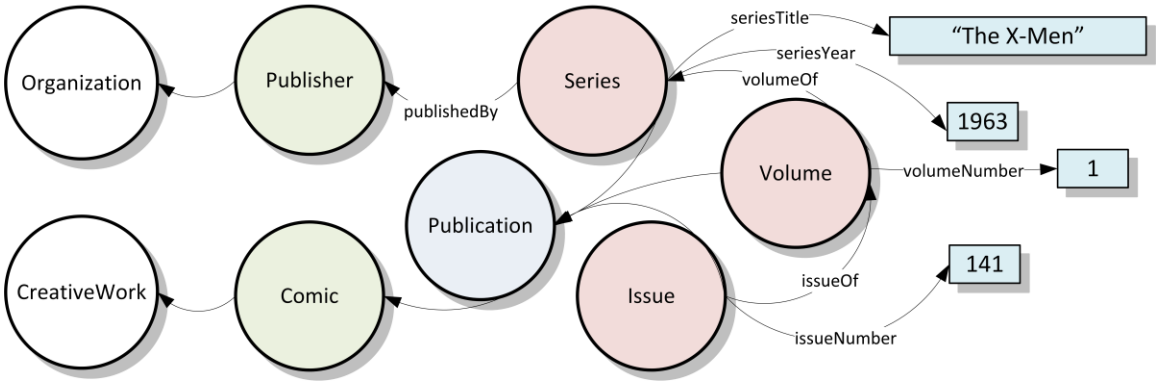

FIG 1. Core concepts in the Comic Book Ontology (CBO).

This poster describes progress on the Comic Book Ontology (CBO) by presenting a diagram illustrating current components of the model (FIG. 1), and outlines the methodology and rationale for producing a core application profile. Additionally, it presents a workflow illustrating how the core set of elements is used to map user data to the vocabulary and generate RDF/XML records through an automated process. Community data is commonly contained in spreadsheets, or made available as CSV, and a workflow is described for both the preparation and conversion of that data, as well as its connection to existing Linked Open Data (LOD) resources.

2. Background

The recent success of Marvel's *Guardians of the Galaxy* at the box-office highlights the dominance of the superhero movie in popular culture, and interest in the genre is only likely to

continue with future films planned featuring familiar icons like Batman, Superman, and Spider-Man. However, before these characters and stories made it to movie screens, they first appeared in periodical comic books on newsstands, where they then made it into the homes and collections of many generations of readers around the world. In addition to appearing in library special collections and archives, like the Comic Art Collection of the Michigan State University Library composed of over 200,000 items (comics.lib.msu.edu), the comic book is also collected by the Library of Congress (LOC) and the institution's Comic Book Collection contains over 120,000 comic issues (LOC, 2013). While the efforts of these institutions are significant, parallel activities occur daily in the homes of many comic book collectors (Serchay, 1998). Passion and dedication to the hobby on the part of both collectors and professionals has produced numerous research projects and efforts dedicated to the comic book. Notable projects in this area include the *Grand Comics Database* (GCD), an international effort to index all comic books published worldwide (comics.org), and *Comichron: The Comic Book Chronicles*, a research project collecting comic book sales and circulation data (comichron.com), among many other related endeavors that can be found in the *Comics Research Bibliography* (Rhode & Bullough, 2009). The Comic Book Ontology (CBO) represents an effort to bring greater bibliographic control, representation, and visibility to the endeavors of many writers, artists, researchers, and collectors who have contributed to the preservation and proliferation of the medium.

3. Application Profile and Workflow

The comic book is a complex object that can be viewed as a bibliographic resource, collection item, and art object, with its contents telling part of the story in an ongoing narrative that can span multiple issues, volumes, and series titles, all of which compose a detailed, fictional universe. In addition to the complexities of the objects themselves, the domain's many participants, including libraries and archives, each produce data of various degrees of quality and control, while following different standards and practices. However, shared entities and elements found in the data formulate a pattern or core model that can represent a simplified view of this complex world.

The methodology for producing the core application profile involved aligning components of the Comic Book Ontology (CBO) to a WEMI model. The WEMI model produces a view of the core elements at various levels of description, up to a specific, physical copy in a comic book collection. Extending the exchange of knowledge to the collector using Linked Data enables a passionate and dedicated segment of the user population to participate in the ecosystem, not just at the item-level, but at all levels of resource description expanding the available information about comic works, creators, and collections. However, in order to participate successfully, users require a simple, clear process for the preparation and conversion of their data. This workflow involves: (1) mapping existing data to CBO terms, (2) converting data to qualified RDF/XML, and (3) automatically replacing values with LOD URIs. The automated conversion process is achieved through an online tool, or an XSLT stylesheet that can be applied locally. Experienced users can modify, rewrite, or create their own script, and expand on the selection of LOD resources linked in the resulting dataset.

4. Summary

The Comic Book Ontology (CBO) seeks to provide the tools through which collectors, researchers, and libraries can share information about their individual collections and better combine and exchange knowledge in a Linked Data environment. In order to improve the usability of the ontology, a core application profile has been developed. A basic workflow describes using this profile to guide the preparation and mapping of existing data to CBO elements, and the automated conversion of that data to qualified RDF/XML containing Linked Data URIs for common values. The core application profile will form the base of additional profiles that will address the needs of other user groups as the ontology develops and additional needs are discovered. The vocabulary is made available at comicmeta.org, which functions as a repository for the ontology as well as all related schemata, tools, and utilities.

References

Library of Congress. (2013, April). Comic book collection. Retrieved from http://www.loc.gov/rr/news/coll/049.html

Rhode, Michael, and John Bullough. (2009). Comics research bibliography. Retrieved from http://homepages.rpi.edu/~bulloj/comxbib.html

Serchay, David. S. (1998). Comic book collectors: The serials librarians of the home. Serials Review, 24(1), 57-70. doi:10.1016/S0098-7913(99)80103-8